Conditional Expected Gain

Robert M. Haralick

Computer Science, Graduate Center
City University of New York

Conditional Expectation

Definition

Let X and Y be a discrete random variables that take values from the set $A \times B$. The Conditional Expectation of Y given X is defined by

$$E[Y \mid X = a] = \sum_{b \in B} b P_{XY}(a, b)$$

$E[Y \mid X]$ is a function of the various values that X can take.

The Event (c^j, c^k, d)

Recall

$$
\begin{aligned}
P_{TA}(c^j, c^k, d) &= P_{TA}(c^j, c^k|d)P(d) \\
&= P_T(c^j|d)P_A(c^k|d)P(d) \\
&= \frac{P_T(d|c^j)P_T(c^j)}{P(d)}P_A(c^k|d)P(d) \\
&= P_T(d|c^j)P_A(c^k|d)P_T(c^j) \\
P_{AT}(c^k, d|c^j) &= \frac{P_{TA}(c^j, c^k, d)}{P_T(c^j)} \\
&= P_T(d|c^j)P_A(c^k|d) = P_T(d|c^j)f_d(c_k)
\end{aligned}
$$

Expected Conditional Economic Gain Given Class

Definition

The conditional expectation of the economic gain given class c^j for decision rule f is defined by

$$
\begin{aligned}
E[e \mid c^j; f] &= \sum_{d \in D} \sum_{k=1}^{K} e(c^j, c^k) P_{TA}(c^j, c^k, d) \\
&= \sum_{d \in D} \sum_{k=1}^{K} e(c^j, c^k) P(d \mid c^j) f_d(c^k) \\
&= \sum_{k=1}^{K} e(c^j, c^k) \sum_{d \in D} P(d \mid c^j) f_d(c^k)
\end{aligned}
$$

where $f_d(c)$ is the conditional probability that the decision rule assigns class c given measurement d.

Class Conditional Probability and Prior Probability

- $P(d|c)$
  - Conditional probability of measurement d given class c
  - Class conditional probability
- $P(c)$
  - Prior probability of class c
  - Prior probability

Economic Gain

The Expected economic gain can be related to the class conditional expected economic gain and prior probabilities.

$$
\begin{aligned}
E[e; f] &= \sum_{d \in D} \sum_{k=1}^{K} \sum_{j=1}^{K} e(c^j, c^k) P(c^j, d) f_d(c^k) \\
&= \sum_{j=1}^{K} \sum_{k=1}^{K} \sum_{d \in D} e(c^j, c^k) P(d \mid c^j) P(c^j) f_d(c^k) \\
&= \sum_{j=1}^{K} \left[\sum_{k=1}^{K} \sum_{d \in D} e(c^j, c^k) P(d \mid c^j) f_d(c^k) \right] P(c^j) \\
&= \sum_{j=1}^{K} E[e \mid c^j; f] P(c^j)
\end{aligned}
$$

Economic Gain

When the economic gain is represented in terms of the prior class probabilities, we write

$$E[e; f, P(c^1), \ldots, P(c^K)]$$

When f is a Bayes decision rule,

$$E[e; f, P(c^1), \ldots, P(c^K)] \geq E[e; g, P(c^1), \ldots, P(c^K)]$$

for any other decision rule g.

Definition

When f is a Bayes decision rule, $E[e; f, P(c^1), \ldots, P(c^K)]$ is called the Bayes gain.

The Geometry of a Bayes Rule

We will show that the geometry of a Bayes Rule is related to convex combinations and convex sets

Convex Combinations

Definition

Let $x, y \in \mathbb{R}^N$ and $0 \leq \lambda \leq 1$. Then $\lambda x + (1-\lambda)y$ is called a convex combination of x and y.

Proposition

If $0 \leq x, y, \lambda \leq 1$, *then* $0 \leq \lambda x + (1-\lambda)y \leq 1$

Proof.

$0 \leq x, y, \lambda$ *implies* $\lambda x + (1-\lambda)y \leq \lambda + (1-\lambda) = 1$.
$\lambda \leq 1$ *implies* $0 \leq 1 - \lambda$.
$x, y, \lambda, 1-\lambda \geq 0$ *implies* $\lambda x + (1-\lambda)y \geq 0$.
Therefore, $0 \leq \lambda x + (1-\lambda)y \leq 1$. □

Structure of Decision Rules

Consider the structure of a decision rule $f_d(c)$.
Suppose $D = \{d^1, \ldots, d^Q\}$ and $C = \{c^1, \ldots, c^K\}$.
Then this decision rule f can be thought of as a vector in $\mathbb{R}^{KQ}$

$$f' = (f_{d^1}(c^1), \ldots, f_{d^1}(c^K), \ldots, f_{d^Q}(c^1), \ldots, f_{d^Q}(c^K))$$

There are some constraints:

- For $q \in \{1, \ldots, Q\}$ and $k \in \{1, \ldots, K\}$, $0 \leq f_{d^q}(c^k) \leq 1$
- For $q \in \{1, \ldots, Q\}$, $\sum_{k=1}^{K} f_{d^q}(c^k) = 1$

Therefore, a decision rule must lie in the unit hypercube of $\mathbb{R}^{QK}$ and it must lie in the manifold defined by the Q linear constraints

$$\sum_{k=1}^{K} f_{d^q}(c^k) = 1, \; q = 1, \ldots, Q$$

8 Possible Deterministic Decision Rules

	d^1	d^2	d^3
f_d^1	c_1	c_1	c_1
f_d^2	c_1	c_1	c_2
f_d^3	c_1	c_2	c_1
f_d^4	c_1	c_2	c_2
f_d^5	c_2	c_1	c_1
f_d^6	c_2	c_1	c_2
f_d^7	c_2	c_2	c_1
f_d^8	c_2	c_2	c_2

Deterministic Decision Rules Written as Probabilistic

$f_d^1(c_1)$ is the probability that decision rule f^1 assigns class c_1 to d^1
$f_d^1(c_2)$ is the probability that decision rule f^1 assigns class c_2 to d^1

f_d^n	$f_{d^1}^n(c_1)$	$f_{d^1}^n(c_2)$	$f_{d^2}^n(c_1)$	$f_{d^2}^n(c_2)$	$f_{d^3}^n(c_1)$	$f_{d^3}^n(c_2)$
f_d^1	1	0	1	0	1	0
f_d^2	1	0	1	0	0	1
f_d^3	1	0	0	1	1	0
f_d^4	1	0	0	1	0	1
f_d^5	0	1	1	0	1	0
f_d^6	0	1	1	0	0	1
f_d^7	0	1	0	1	1	0
f_d^8	0	1	0	1	0	1

$$f_{d^n}(c^1) + f_{d^n}(c^2) = 1, \qquad n = 1, 2, 3$$
$$0 \leq f_{d^n}(c^k) \leq 1, \qquad n = 1, 2, 3; \; k = 1, 2$$

Deterministic Decision Rule Written Probabilisticly

$$f_{d^n}(c^2) = 1 - f_{d^n}(c^1), \qquad n = 1, 2, 3$$
$$0 \leq f_{d^n}(c^1) \leq 1, \qquad n = 1, 2, 3$$

$f_d^n(c^1)$	d^1	d^2	d^3
f_d^1	1	1	1
f_d^2	1	1	0
f_d^3	1	0	1
f_d^4	1	0	0
f_d^5	0	1	1
f_d^6	0	1	0
f_d^7	0	0	1
f_d^8	0	0	0

Mixture Decision Rules

Let $0 \leq \lambda \leq 1$ What does $g_d = \lambda f_d^2 + (1-\lambda) f_d^7$ mean?
With probability λ choose decision rule f_d^2 and with probability $1-\lambda$ choose decision rule f_d^7

	d^1	d^2	d^3
$g_d(c^1)$	λ	λ	$1-\lambda$
$g_d(c^2)$	$1-\lambda$	$1-\lambda$	λ

$$F = \{ f_d(c^1) \mid f_d(c^1) = \sum_{n=1}^{8} \lambda_n f_d^n(c^1), \text{ for some } 0 \leq \lambda_n \leq 1, \sum_{n=1}^{8} \lambda_n = 1 \}$$

F is the set of all convex combinations of the decision rules $f_d^1, \ldots, f_d^8$.
The convex combinations are probabilistic decision rules.

Convex Combinations of Probabilistic Decision Rules

Proposition

Convex combinations of decision rules are decision rules

Proof.

Let f and g be two decision rules. Let $0 \leq \lambda \leq 1$. *Consider* $\lambda f_d(c) + (1 - \lambda) g_d(c)$. *We have already proven that* $0 \leq \lambda f_d(c) + (1 - \lambda) g_d(c) \leq 1$. *Consider the convex combination:*

$$\begin{aligned} \sum_{c \in C} [\lambda f_d(c) + (1 - \lambda) g_d(c)] &= \lambda \sum_{c \in C} f_d(c) + (1 - \lambda) \sum_{c \in C} g_d(c) \\ &= \lambda + (1 - \lambda) \\ &= 1 \end{aligned}$$

□

Convex Sets

Definition

A set $C \subseteq \mathbb{R}^N$ is a convex set if and only if $x, y \in C$ imply $\lambda x + (1 - \lambda)y \in C$ for every $0 \leq \lambda \leq 1$.

Proposition

The set F of all convex combinations of decision rules is a convex set.

Example

$$F = \{f_d(c^1) \mid f_d(c^1) = \sum_{n=1}^{8} \lambda_n f_d^n(c^1), \text{ for some } 0 \leq \lambda_n \leq 1, \sum_{n=1}^{8} \lambda_n = 1\}$$

Intersection of Convex Sets are Convex

Proposition

Let C and D be convex sets. Then $C \cap D$ is a convex set.

Proof.

Let $x, y \in C \cap D$ and $0 \leq \lambda \leq 1$. Consider $\lambda x + (1 - \lambda)y$.
Since $x, y \in C \cap D$, $x, y \in C$ and $x, y \in D$.
Since C is convex and $0 \leq \lambda \leq 1$, $\lambda x + (1 - \lambda)y \in C$.
Since D is convex and $0 \leq \lambda \leq 1$, $\lambda x + (1 - \lambda)y \in D$.
$\lambda x + (1 - \lambda)y \in C$ and $\lambda x + (1 - \lambda)y \in D$ imply
$\lambda x + (1 - \lambda)y \in C \cap D$. □

Mixed Decision Rules

Definition

Let f and g be decision rules and $0 \leq \lambda \leq 1$.
Then

$$h_d(c) = \lambda f_d(c) + (1 - \lambda) g_d(c)$$

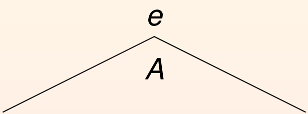
is called a mixed decision rule of f and g.

- With probability λ apply decision rule f and probability $1 - \lambda$ apply decision rule g.
- If we apply decision rule f, the we assign class c with probability $f(c|d)$
- If we apply decision rule g, then we assign class c with probability $g(c|d)$

Extreme Points

Definition

Let $A \subseteq \mathbb{R}^N$. A point $e \in A$ is called an Extreme Point of A if and only if $b, c \in A$ with $e = \frac{b+c}{2}$ implies $e = b = c$.



If e is an extreme point of A and if $b,\ c \in A$ and for some λ, $0 \leq \lambda \leq 1$ then

$$e = \lambda b + (1 - \lambda)c \text{ implies } e = b = c$$

If e is an extreme point of A then there is no convex combination of a distinct pair of points in A that equals e.

Deterministic Decision Rules are Extreme Points

Proposition

Let F be the set of all convex combinations of decision rules. Let f be a deterministic decision rule. Then f is an extreme point of F.

Proof.

Let $g, h \in F$ satisfy $f = \frac{g+h}{2}$. Hence for every $d \in D$ and $c \in C$,

$$f_d(c) = \frac{g_d(c) + h_d(c)}{2}$$

Since f is a deterministic decision rule, for some $c^* \in C$, $f_d(c^*) = 1$ and for all $c \in C - \{c^*\}$, $f_d(c) = 0$. Consider $c \in C$ for which $f_d(c) = 0$.

$$f_d(c) = 0 = \frac{g_d(c) + h_d(c)}{2}$$

Since $g_d(c), h_d(c) \geq 0$ and since $g_d(c) + h_d(c) = 0$, it follows that $g_d(c) = h_d(c) = 0$. □

Proof Continued

Proof.

Now consider c^*.

$$f_d(c^*) = 1 = \frac{g_d(c^*) + h_d(c^*)}{2}$$

Hence, $g_d(c^*) + h_d(c^*) = 2$. But $g_d(c^*), h_d(c^*) \leq 1$. Therefore, $g_d(c^*) = 1$ and $h_d(c^*) = 1$.

Now, by definition of extreme point, a deterministic decision rule $f \in F$ is an extreme point of F, the set of all convex combinations of decision rules. $\square$

Convex Polyhedrons

Definition

A Closed Convex Polyhedron is a non-empty set P formed as the solutions to a matrix equation $Ax \leq b$.

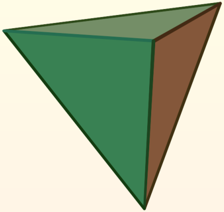
$$P = \{x \mid Ax \leq b\}$$

Each row of the matrix equation specifies a hyperplane half space and P is the intersection of these hyperplane half spaces.

Definition

A bounded polyhedron is a polytope.

Closed Convex Polytope Example Tetrahedron



$$P = \{x \mid Ax \leq b\}$$

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & -1 \\ -1 & 1 & -1 \\ -1 & -1 & 1 \end{pmatrix}$$

$$b = \begin{pmatrix} 2 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

The Set of Decision Rules is a Closed Convex Polytope

Proposition

Let F be the set of all decision rules formed from the finite set C of classes and the finite set D of measurements.
The set F is a closed convex polytope lying in a linear manifold of dimension $|C|\,|D| - |D|$.

Proof.

Let $f \in F$. We already know that $f \in \mathbb{R}^{|C|\,|D|}$. The $|D|$ linear constraints are formed from the requirement that $\sum_{c \in C} f_d(c) = 1$. The remaining constraints are of the form

- $f_d(c) \geq 0$ *which is equivalent to* $-f_d(c) \leq 0$
- $f_d(c) \leq 1$

□

Minkowski's Theorem

Definition

Let $X = \{x_1, \ldots, x_M\} \subset \mathbb{R}^N$. The Convex Hull of X is defined by

$$\mathcal{CH}(X) = \{y \in \mathbb{R}^N \mid y = \sum_{m=1}^{M} \lambda_m x_m, \text{where } \lambda_m \geq 0, \sum_{m=1}^{M} \lambda_m = 1\}$$

Theorem

Any closed convex polytope is the convex hull of its extreme points.

Probabilistic Decision Rules

Any Probabilistic Decision Rule can be represented as a convex combination of the deterministic decision rules.

Theorem

Let f be a probabilistic decision rule and let $f^1, \ldots, f^M$ be the set of all possible deterministic decision rules. Then there exists a convex combination $\lambda_1, \ldots, \lambda_M$ such that

$$f_d(c) = \sum_{m=1}^{M} \lambda_m f_d^m(c)$$

Proposition

Let $C \subseteq \mathbb{R}^N$ be a convex set. Let e be an extreme point of C. Let D be a convex subset of C. If $e \in D$, then e is an extreme point of D.

Proof.

Let e be an extreme point of C. Suppose $e \in D$. Let $a, b \in D$ satisfy $e = \frac{a+b}{2}$. Since $D \subseteq C$, $a, b \in C$. Now, $a, b \in D \subseteq C$, with $e = \frac{a+b}{2}$. Since e is an extreme point of C, $e = a = b$. But now we have $e \in D$ and $a, b \in D$ satisfying $e = \frac{a+b}{2}$. And we have just proved that $e = a = b$. Therefore, e is an extreme point of D. $\square$

Expected Conditional Gain: Mixed Decision Rules

Proposition

$$E[e \mid c^j; \lambda f + (1-\lambda)g] = \lambda E[e \mid c^j; f] + (1-\lambda)E[e \mid c^j; g]$$

Proof.

$$\begin{aligned}
E[e \mid c^j; \lambda f + (1-\lambda)g] &= \sum_{k=1}^{K}\sum_{d\in D} e(c^j, c^k)P(d \mid c^j)\{\lambda f(c^k \mid d) + (1-\lambda)g(c^k \mid d)\} \\
&= \lambda \sum_{k=1}^{K}\sum_{d\in D} e(c^j, c^k)P(d \mid c^j)f(c^k \mid d) + \\
&\quad (1-\lambda)\sum_{k=1}^{K}\sum_{d\in D} e(c^j, c^k)P(d \mid c^j)g(c^k \mid d) \\
&= \lambda E[e \mid c^j; f] + (1-\lambda)E[e \mid c^j; g]
\end{aligned}$$

□

Example

e	Assigned	
True	c^1	c^2
c^1	2	-1
c^2	-1	2

$P(d \mid c)$	Measurement		
True Class	d^1	d^2	d^3
c^1	.2	.3	.5
c^2	.5	.4	.1

$f_d(c)$	Measurement		
True Class	d^1	d^2	d^3
c^1	1	0	0
c^2	0	1	1

$$E[e \mid c^j; f] \quad = \quad \sum_{d \in D} \sum_{k=1}^{K} e(c^j, c^k) P(d \mid c^j) f_d(c^k)$$

$$\begin{aligned}
E[e \mid c^1; f] &= e(c^1, c^1)P(d^1 \mid c^1)f_{d^1}(c^1) + e(c^1, c^2)P(d^1 \mid c^1)f_{d^1}(c^2) + \\
&\quad e(c^1, c^1)P(d^2 \mid c^1)f_{d^2}(c^1) + e(c^1, c^2)P(d^2 \mid c^1)f_{d^2}(c^2 + \\
&\quad e(c^1, c^1)P(d^3 \mid c^1)f_{d^3}(c^1) + e(c^1, c^2)P(d^3 \mid c^1)f_{d^3}(c^2) \\
&= 2 * .2 * 1 + (-1) * .2 * 0 + \\
&\quad 2 * .3 * 0 + (-1) * .3 * 1 + \\
&\quad 2 * .5 * 0 + (-1) * .5 * 1 \\
&= .4 - .3 - .5 = -.4
\end{aligned}$$

Example

e	Assigned	
True	c^1	c^2
c^1	2	-1
c^2	-1	2

$P(d \mid c)$	Measurement		
True Class	d^1	d^2	d^3
c^1	.2	.3	.5
c^2	.5	.4	.1

$f_d(c)$	Measurement		
True Class	d^1	d^2	d^3
c^1	1	0	0
c^2	0	1	1

$$E[e \mid c^j; f] = \sum_{d \in D} \sum_{k=1}^{K} e(c^j, c^k) P(d \mid c^j) f_d(c^k)$$

$$\begin{aligned}
E[e \mid c^2; f] &= e(c^2, c^1)P(d^1 \mid c^2)f_{d^1}(c^1) + e(c^2, c^2)P(d^1 \mid c^2)f_{d^1}(c^2) + \\
&\quad e(c^2, c^1)P(d^2 \mid c^2)f_{d^2}(c^1) + e(c^2, c^2)P(d^2 \mid c^2)f_{d^2}(c^2) + \\
&\quad e(c^2, c^1)P(d^3 \mid c^2)f_{d^3}(c^1) + e(c^2, c^2)P(d^3 \mid c^2)f_{d^3}(c^2) \\
&= (-1) * .5 * 1 + 2 * .5 * 0 + \\
&\quad (-1) * .4 * 0 + 2 * .4 * 1 + \\
&\quad (-1) * .1 * 0 + 2 * .1 * 1 \\
&= -.5 + .8 + .2 = .5
\end{aligned}$$

Example

e	Assigned	
True	c^1	c^2
c^1	2	-1
c^2	-1	2

$P(d \mid c)$	Measurement		
True Class	d^1	d^2	d^3
c^1	.2	.3	.5
c^2	.5	.4	.1

$$E[e \mid c^j; f] = \sum_{d \in D} \sum_{k=1}^{K} e(c^j, c^k) P(d \mid c^j) f_d(c^k)$$

	Measurements			Conditional Gain	
f	d^1	d^2	d^3	$E[e \vert c^1; f]$	$E[e \vert c^2; f]$
f^1	c^1	c^1	c^1	2.0	-1.0
f^2	c^1	c^1	c^2	.5	-.7
f^3	c^1	c^2	c^1	1.1	.2
f^4	c^1	c^2	c^2	-.4	.5
f^5	c^2	c^1	c^1	1.4	.5
f^6	c^2	c^1	c^2	-.1	.8
f^7	c^2	c^2	c^1	.5	1.7
f^8	c^2	c^2	c^2	-1.	2.0

Conditional Expected Gains: All Decision Rules